USDA FOREST SERVICE RESEARCH NOTE NE-236
1977

Northeastern Forest Experiment Station

FOREST SERVICE, U.S. DEPT. OF AGRICULTURE, 6816 MARKET STREET, UPPER DARBY, PA. 19082

ODC [853:525.1]--015.5

RIDGE: A COMPUTER PROGRAM FOR CALCULATING RIDGE REGRESSION ESTIMATES

by **DONALD E. HILT**
Research Forester
and
DONALD W. SEEGRIST
Mathematical Statistician
USDA Forest Service
Northeastern Forest Experiment Station

Abstract:—Least-squares coefficients for multiple-regression models may be unstable when the independent variables are highly correlated. Ridge regression is a biased estimation procedure that produces stable estimates of the coefficients. Ridge regression is discussed, and a computer program for calculating the ridge coefficients is presented.

KEYWORDS: Regression, computer program, correlated variables.

Multiple-regression models are widely used in forestry. In some studies, the independent variables are highly correlated. In this case the least-squares coefficients may be too large in absolute value, and the signs may reverse with small changes in the data. With highly correlated data, one should consider estimation methods that reduce the effects of the correlation and produce stable regression coefficients *(Marquardt and Snee 1975).*

The purpose of this note is to discuss ridge-regression methods and present a computer program for ridge regression. A list of references is also given.

Ridge Regression

The observational equations for a multiple-regression model can be written as

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\epsilon}$$

in which $\underset{\sim}{Y}$ is the nxl vector of observations, $\underset{\sim}{X}$ is the nxp matrix of independent variables, $\underset{\sim}{\beta}$ is a pxl vector of parameters unknown, and $\underset{\sim}{\epsilon}$ is the nxl vector of errors. It is assumed that $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$ and $E(\underset{\sim}{\epsilon}'\underset{\sim}{\epsilon}) = \delta^2\underset{\sim}{I}$.

The least-squares estimate of $\underset{\sim}{\beta}$ is

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}. \qquad [1]$$

For convenience, we assume that $X'X$ and $X'Y$ are in the correlation form. Methods of scaling $X'X$ and $X'Y$ to the correlation form are discussed by Draper and Smith *(1966, p. 147)*. It is well known that $\hat{\beta}$ is the best linear unbiased estimate of β. However, when the predictor variables are highly correlated, the average distance of $\hat{\beta}$ to β is large. In particular, $E[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)]$ is large.

Hoerl and Kennard *(1970a)* suggested that the estimator

$$\hat{\beta}^* = (X'X+kI)^{-1}X'Y;\ k \geq 0 \qquad [2]$$

be used when the independent variables are highly correlated. The estimate $\hat{\beta}^*$ is called the ridge estimator. If $\beta'\beta$ is bounded, there exists a value of $k>0$ such that $E[(\hat{\beta}^*-\beta)'(\hat{\beta}^*-\beta)] < E[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)]$. The ridge estimator has the property that, as k increases, the variance of $\hat{\beta}^*$ decreases, but the bias increases. The best regression estimates of $\hat{\beta}^*$ are those that are stable and have a small mean-square error.

To calculate the ridge estimator $\hat{\beta}^*$ from equation [2], one would have to invert the pxp matrix, $(X'X+kI)$, for each value of k. This sequence of matrix inversions could be time-consuming even with a high-speed computer. The ridge estimator can be expressed in a form that may be better for computing purposes.

We know from matrix theory that, because $X'X$ is symmetric, there exists an orthogonal matrix A and a diagonal matrix D such that $A'X'XA=D$ and $A'A=I$. The matrix A is the matrix of eigenvectors of $X'X$, and the matrix D is the diagonal matrix of eigenvalues of $X'X$. Adding kI to both sides of $A'X'XA=D$ gives

$$A'X'XA+kI=D+kI. \qquad [3]$$

Multiplying the second term on the left-hand side of equation [3] by $A'A$, gives

$$A'X'XA+kA'IA=D+kI, \qquad [4]$$

which can be written as

$$A'(X'X+kI)A=D+kI. \qquad [5]$$

Premultiplying both sides of equation [5] by $(A')^{-1}$ and postmultiplying by A^{-1} gives

$$X'X+kI=(A')^{-1}(D+kI)A^{-1}. \qquad [6]$$

Taking the inverse of both sides yields

$$(X'X+kI)^{-1}=A(D+kI)^{-1}A'. \qquad [7]$$

Substituting the results of equation [7] in equation [2], we find that the ridge estimator can be written

$$\hat{\beta}^*=A(D+kI)^{-1}A'X'Y. \qquad [8]$$

This form of the ridge estimator may be efficient for computing in problems with a large number of independent variables. The matrix $(D+kI)$ is diagonal, and the elements of the inverse are the reciprocals of the diagonal elements. The matrix of eigenvectors A and the matrix of eigenvalues D need to be calculated only once. However, the algorithm for computing the eigenvalues is iterative, and the solution may occasionally take more time than calculating the inverses of $(X'X+kI)$.

The estimates of the ridge coefficients at $k=0$ are the least squares estimates. If the least squares regression is significant, then different values of k should be explored.

The ridge trace, which is a plot of the ridge coefficients for different values of k, is an important part of ridge regression. The sums of squares of residuals should also be plotted. The ridge trace is examined for trends of the ridge coefficients as k is changed. The best estimates of the ridge coefficients are those where the trace shows that the coefficients have stabilized and the sums of squares of residuals is still small *(Marquardt and Snee 1975)*.

Hoerl and Kennard *(1970b)* discuss the use of the ridge trace to eliminate variables with the least predicting power. Thus, ridge regression can be used as a guide for selecting the best subset of variables; that is, ridge regression is an alternative for stepwise regression.

Program Ridge

Program RIDGE is written in ASA Fortran IV for the IBM 370/168 computer. Information needed for the control cards is listed in the appendix. A variable format statement is used to input the data. The dependent variable is positioned by the program, hence special arrangement of the data is not necessary. A maximum of 19 independent variables is allowed for program RIDGE. This capacity may

be increased by changing the dimension statements. Nineteen values of k from 0 to 1.0 are automatically supplied by the program. Other values of k may be designated by the user.

The means and variances of the variables are printed by program RIDGE. The $\underset{\sim}{X}'\underset{\sim}{X}$ and $\underset{\sim}{X}'\underset{\sim}{Y}$ matrices are transformed into the correlation form and printed.

The eigenvalues and corresponding matrix of eigenvectors for the $\underset{\sim}{X}'\underset{\sim}{X}$ matrix are calculated. The presence of one or more zero eigenvalues indicates linear dependencies between the independent variables. If this condition exists, $\underset{\sim}{X}'\underset{\sim}{X}$ is singular for k=0, and the program terminates with an error message. If no linear dependencies are present, an analysis of variance table is printed.

Standardized and actual regression coefficients are printed for the different values of k. The ridge trace can be plotted by the user from the standardized coefficients. However, we found that in most cases the tabled values of standardized coefficients provide sufficient information for selecting the appropriate ridge solution.

The computer program is available from the Biometrics Group, Northeastern Forest Experiment Station.

An Example of Ridge Regression

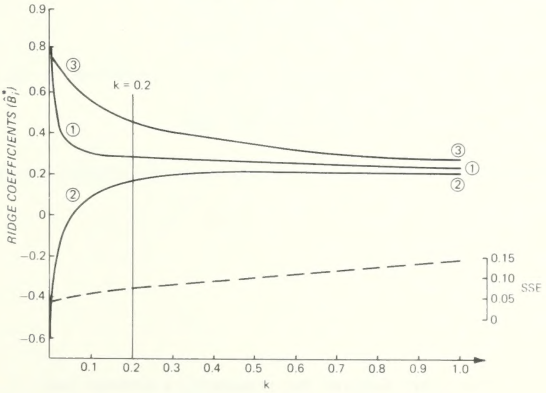
Suppose we have 10 sample observations for 3 independent variables and 1 dependent variable (table 1). Computer output from program RIDGE for this example is given in the appendix. Investigation of the correlation matrix reveals high correlations between the predictor variables; and one of the eigenvalues, 0.0138, is small. These conditions suggest that ridge regression be used to estimate the regression coefficients. Since the F ratio for the least-squares solution is highly significant, ridge-regression coefficients and the residual sums of squares were calculated for 19 values of k.

The ridge trace was constructed by plotting

Table 1.—*Data for sample problem*

Y	X_1	X_2	X_3
223	11	11	11
223	14	15	11
292	17	18	20
270	17	17	18
285	18	19	18
304	18	18	19
311	19	18	20
314	20	21	21
328	23	24	25
340	25	25	24

Figure 1.—Ridge trace for 19 values of k.



the standardized regression coefficients against values of k (fig. 1). The trace suggests that the least-squares coefficients are too large in absolute value, $\hat{\beta}_2$ even having the wrong sign. At k=0.2 the coefficients have stabilized, and the residual sums of squares (SSE) has not substantially increased.

The ridge regression

$$\hat{Y}^* = 132.5 + 2.870(X1) + 1.650(X2) + 3.934(X3)$$

should be a better predicting equation than the least-squares equation even though the coefficients are biased.

Summary

Ridge regression is a statistical technique that foresters should find useful. It is used to estimate coefficients for multiple-regression models when the independent variables are highly correlated.

Considerable research has been done on ridge regression. The paper by Hoerl and Kennard *(1970a)* introduced ridge-regression theory. Although there is considerable matrix algebra in this paper, it provides a sound background for the understanding and application of ridge regression. The subsequent paper by Hoerl and Kennard *(1970b)* illustrates the applications of ridge regression, including its use as a guide to variable selection. The article by Marquardt and Snee *(1975)* is perhaps the most readable paper on ridge regression. All aspects of ridge regression are discussed at length, and many examples are included. Some of the other articles listed are more mathematically sophisticated.

Literature References

Brown, P., and C. Payne.
1975. **Election night forecasting.** J. Roy. Statist. Soc. Ser. A. 38:463-498.

Conniffe, D., and Joan Stone.
1974. **A critical review of RIDGE regression.** Statistician 22:181-187.

Draper, N. R., and H. Smith.
1966. **Applied regression analysis.** Wiley, New York. 407 p.

Farebrother, R. W.
1975. **The minimum mean square error linear estimator and RIDGE regression.** Technometrics 17:127-128.

Giulkey, David K., and James L. Murphy.
1975. **Directed RIDGE regression techniques in cases of multicollinearity.** J. Am. Statist. Assoc. 70:769-775.

Hemmerle, William J.
1975. **An explicit solution for generalized RIDGE regression.** Technometrics 17:309-314.

Hoerl, Arthur E., and Robert W. Kennard.
1970a. **RIDGE regression: applications to nonorthogonal problems.** Technometrics 12:55-68.

Hoerl, Arthur E., and Robert W. Kennard.
1970b. **RIDGE regression: biased estimation for nonorthogonal problems.** Technometrics 12:69-83.

Marquardt, Donald W.
1970. **Generalized inverses, RIDGE regression, biased linear estimation, and nonlinear estimation.** Technometrics 12:591-612.

Marquardt, Donald W., and Ronald D. Snee.
1975. **RIDGE regression in practice.** Am. Statist. 29:3-19.

MANUSCRIPT RECEIVED FOR PUBLICATION 6 OCTOBER 1976

APPENDIX

```
      PROGRAM RIDGE REGRESSION


CXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXC
C                                                                                C
C                                                                                C
C     PROGRAM CONTROL INFORMATION:                                               C
C                                                                                C
C     PROGRAM CONTROL CARDS MUST BE THE FIRST CARDS IN THE DATA DECK.           C
C                                                                                C
C          CARD 1 (REQUIRED): PROBLEM TITLE,UP TO 80 CHARACTERS LONG.           C
C                             A BLANK CARD MAY BE SUBMITTED IF NO                C
C                             PROBLEM TITLE IS DESIRED.                          C
C                                                                                C
C          CARD 2 (REQUIRED): SPECIFY N,M,PY,D,AND V. FORMAT IS 5I5,            C
C                             RIGHT JUSTIFIED.                                   C
C                                                                                C
C                             N = NUMBER OF OBSERVATIONS.                        C
C                             M = NUMBER OF VARIABLES,INCLUDING Y.               C
C                                 MAXIMUM OF 19 INDEPENDENT VARIABLES.           C
C                             PY = POSITION OF THE DEPENDENT VARIABLE            C
C                                  IN THE DATA. THE PROGRAM WILL MAKE            C
C                                  THE DEPENDENT VARIABLE THE LAST               C
C                                  VARIABLE.                                     C
C                             D = NUMBER OF INCREMENTS (K'S) FOR THE             C
C                                 X-PRIME-X MATRIX, IF INCREMENTS ARE            C
C                                 TO BE USER-SUPPLIED. MAXIMUM OF 18             C
C                                 K'S MAY BE SPECIFIED. INCREMENTS               C
C                                 WILL BE PROGRAM-SUPPLIED IF LEFT               C
C                                 BLANK. K = 0.0 IS ALWAYS SUPPLIED              C
C                                 BY THE PROGRAM, AND SHOULD NOT BE              C
C                                 SPECIFIED BY THE USER. CARD 5                  C
C                                 REQUIRED IF D IS NOT BLANK.                    C
C                             V = 1 IF VARIABLE NAMES ARE TO BE                  C
C                                 SPECIFIED BY THE USER. LEAVE BLANK             C
C                                 OTHERWISE. CARD 4 REQUIRED IF V IS             C
C                                 NOT BLANK.                                     C
C                                                                                C
C          CARD 3 (REQUIRED): VARIABLE FORMAT FOR DATA, ENCLOSED IN              C
C                             PARENTHESES.                                       C
C                                                                                C
C          CARD 4 (OPTIONAL): VARIABLE NAMES. FORMAT IS MA8, LEFT               C
C                             JUSTIFIED. BLANKS MUST BE LEFT FOR THOSE           C
C                             VARIABLES WITH NO NAMES IF THIS OPTION             C
C                             IS IN EFFECT. MORE THAN ONE CARD IF                C
C                             NECESSARY.                                         C
C                                                                                C
C          CARD 5 (OPTIONAL): VALUES OF INCREMENTS,K,TO BE ADDED TO             C
C                             THE X-PRIME-X MATRIX. FORMAT IS DF5.0.            C
C                             VALUES SHOULD HAVE DECIMAL POINTS.                C
C                             MORE THAN ONE CARD IF NECESSARY.                  C
C                                                                                C
C                                                                                C
CXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXC
```

RIDGE REGRESSION
NORTHEASTERN FOREST EXPERIMENT STATION
UPPER DARBY, PA.
JUNE 4, 1976

VARIABLE NO.	NAME	MEAN	VARIANCE
1	X1	0.18200E+02	0.16178E+02
2	X2	0.18600E+02	0.16711E+02
3	X3	0.18700E+02	0.21789E+02
4	Y	0.28900E+03	0.16171E+04

CORRELATION MATRIX R

VARIABLE NO.	NAME				
1	X1	1.0000			
2	X2	0.9853	1.0000		
3	X3	0.9386	0.9247	1.0000	
4	Y	0.9384	0.9064	0.9725	1.0000

EIGENVALUES OF X-PRIME-X

VARIABLE NO.	NAME	
1	X1	2.8993
2	X2	0.0869
3	X3	0.0138

MATRIX OF EIGENVECTORS OF X-PRIME-X

0.5824	-0.3219	-0.7465
0.5796	-0.4795	0.6589
0.5700	0.8164	0.0927

ANALYSIS OF VARIANCE TABLE FOR
LEAST SQUARES SOLUTION (K=0.0)

SOURCE	DF	SUMS OF SQUARES	MSE	F
TOTAL	9	0.14554E+05		
REGRESSION	3	0.14008E+05	0.46694E+04	0.51332E+02
RESIDUAL	6	0.54579E+03	0.90965E+02	

RIDGE TRACE FOR STANDARDIZED REGRESSION COEFFICIENTS

VALUE OF K

VARIABLE NAME							
	0.0	0.020	0.040	0.060	0.080	0.100	0.150	0.200
	0.250	0.300	0.350	0.400	0.450	0.500	0.600	0.700
	0.800	0.900	1.000					
B 1 X1	0.82425E+00	0.44130E+00	0.35900E+00	0.32723E+00	0.31181E+00	0.30315E+00	0.29252E+00	0.28709E+00
	0.28309E+00	0.27955E+00	0.27616E+00	0.27283E+00	0.26952E+00	0.26623E+00	0.25974E+00	0.25340E+00
	0.24724E+00	0.24129E+00	0.23556E+00					
B 2 X2	-.61820E+00	-.19054E+00	-.57015E-01	0.13927E-01	0.59237E-01	0.90985E-01	0.14012E+00	0.16773E+00
	0.18481E+00	0.19591E+00	0.20331E+00	0.20826E+00	0.21151E+00	0.21355E+00	0.21520E+00	0.21478E+00
	0.21313E+00	0.21074E+00	0.20787E+00					
B 3 X3	0.77053E+00	0.71965E+00	0.66183E+00	0.61559E+00	0.57879E+00	0.54897E+00	0.49428E+00	0.45665E+00
	0.42875E+00	0.40692E+00	0.38914E+00	0.37421E+00	0.36137E+00	0.35013E+00	0.33110E+00	0.31535E+00
	0.30189E+00	0.29011E+00	0.27961E+00					
SSE	0.038	0.043	0.048	0.052	0.056	0.059	0.065	0.071
	0.075	0.080	0.084	0.088	0.092	0.097	0.105	0.114
	0.123	0.132	0.142					

ACTUAL BETAS CORRESPONDING TO STANDARDIZED COEFFICIENTS

VALUE OF K

VARIABLE NAME								
	0.0	0.020	0.040	0.060	0.080	0.100	0.150	0.200
	0.250	0.300	0.350	0.400	0.450	0.500	0.600	0.700
	0.800	0.900	1.000					
B 0	0.12800E+03	0.12754E+03	0.12749E+03	0.12774E+03	0.12818E+03	0.12875E+03	0.13051E+03	0.13250E+03
	0.13460E+03	0.13673E+03	0.13886E+03	0.14097E+03	0.14304E+03	0.14508E+03	0.14902E+03	0.15279E+03
	0.15638E+03	0.15980E+03	0.16306E+03					
B 1 X1	0.82408E+01	0.44171E+01	0.35892E+01	0.32717E+01	0.31175E+01	0.30309E+01	0.29246E+01	0.28703E+01
	0.28303E+01	0.27949E+01	0.27610E+01	0.27277E+01	0.26946E+01	0.26617E+01	0.25968E+01	0.25334E+01
	0.24719E+01	0.24124E+01	0.23551E+01					
B 2 X2	-.60813E+01	-.18743E+01	-.56086E+00	0.13700E+00	0.58272E+00	0.89502E+00	0.13784E+01	0.16500E+01
	0.18180E+01	0.19272E+01	0.20000E+01	0.20486E+01	0.20806E+01	0.21007E+01	0.21169E+01	0.21128E+01
	0.20966E+01	0.20730E+01	0.20449E+01					
B 3 X3	0.66381E+01	0.61998E+01	0.57016E+01	0.53032E+01	0.49863E+01	0.47293E+01	0.42582E+01	0.39340E+01
	0.36937E+01	0.35056E+01	0.33524E+01	0.32238E+01	0.31132E+01	0.30163E+01	0.28524E+01	0.27167E+01
	0.26008E+01	0.24992E+01	0.24088E+01					